arXiv:1203.1958v1 [math.PR] 8 Mar 2012

# NUMERICAL COMPUTATION OF CONVOLUTIONS IN FREE PROBABILITY THEORY

SHEEHAN OLVER AND RAJ RAO NADAKUDITI

Abstract. We develop a numerical approach for computing the additive, multiplicative and compressive convolution operations from free probability theory. We utilize the regularity properties of free convolution to identify (pairs of) 'admissible' measures whose convolution results in a so-called 'invertible measure' which is either a smoothly-decaying measure supported on the entire real line (such as the Gaussian) or square-root decaying measure supported on a compact interval (such as the semi-circle). This class of measures is important because these measures along with their Cauchy transforms can be accurately represented via a Fourier or Chebyshev series expansion, respectively. Thus knowledge of the functional inverse of their Cauchy transform suffices for numerically recovering the invertible measure via a non-standard yet well-behaved Vandermonde system of equations. We describe explicit algorithms for computing the inverse Cauchy transform alluded to and recovering the associated measure with spectral accuracy.

## 1. Introduction

We propose a powerful method that allows us to numerically calculate the 'free' [24] additive, multiplicative and compressive convolution of a large class of probability measures. We see this method as complementing the symbolic techniques previously developed in [19] for so-called algebraic measures, *i.e.*, measures whose Cauchy transform is algebraic.

Using the method developed in this paper, we can, for example, compute *with spectral accuracy* the free additive convolution of the semi-circle and the Gaussian which arises in [8] (see Figure 2); or the free compression of the Gaussian which arises in [2] (see Figure 9); or even the free additive convolution of the Gaussian with the counting measure on a single realization of a Gaussian Orthogonal Ensemble as a way to get insight on the rate of convergence to the asymptotic result in [8] (see Figures 2 & 3). We go well beyond these simple examples and hope that the proposed method allows practitioners to experiment with these convolutions in the context of their experiments so they may find new applications of the underlying theory.

We consider the free convolution operations on measures $\mu_A$ and $\mu_B$ (and compression factor $\alpha \in (0,1)$) listed in the first column of Table 1. Each operation takes in one or two measures, and returns a new measure. What is known in each case is a relationship in transform space; i.e., there are transforms [21, 22, 24, 14] $R_\mu(y)$ and $S_\mu(y)$ so that the convolution operation can be represented simply as in the second column of Table 1.

*Date*: February 25, 2013.

2000 *Mathematics Subject Classification.* 15A52, 46L54, 60F99.

*Key words and phrases.* Random matrices, free probability.

| **Operation** | **Transform Operation** | **Key Transform** |
|---|---|---|
| Free Addition<br>$\mu_C = \mu_A \boxplus \mu_B$ | $R_{\mu_C}(y) = R_{\mu_A}(y) + R_{\mu_B}(y)$ | $G_\mu(z) = \int \frac{\mathrm{d}\mu(x)}{z-x}$,<br>$R_\mu(y) = G_\mu^{-1}(y) - \frac{1}{y}$ |
| Free Multiplication<br>$\mu_C = \mu_A \boxtimes \mu_B$ | $S_{\mu_C}(y) = S_{\mu_A}(y)\, S_{\mu_B}(y)$ | $T_\mu(z) = \int \frac{x\mathrm{d}\mu(x)}{z-x}$,<br>$S_\mu(y) = \frac{1+y}{y} T_\mu^{-1}(y)$ |
| Free Compression<br>$\mu_C = \alpha \boxdot \mu_A$ | $R_{\mu_C}(y) = R_{\mu_A}(\alpha y)$ | $G_\mu(z) = \int \frac{\mathrm{d}\mu(x)}{z-x}$,<br>$R_\mu(y) = G_\mu^{-1}(y) - \frac{1}{y}$ |

Table 1. Free convolution operations considered in this paper.

The *Cauchy transform* of a measure $\mu$ on the real line is defined as:

$$G_\mu(z) = \int \frac{\mathrm{d}\mu(x)}{z-x} \qquad \text{for } z \notin \operatorname{supp} \mu.$$

The key observation is that each transform $R_\mu$ and $S_\mu$ can be expressed in terms of the functional inverse of the Cauchy transform $G_\mu^{-1}$ (which we refer to as the *inverse Cauchy transform*). Therefore, we reduce the problem to the following two subtasks:

(1) calculate the inverse Cauchy transform of the input measures pointwise; and
(2) recover the output measure from knowledge of its inverse Cauchy transform.

1.1. **Invertible measures, their utility and the key underlying idea.** In this paper, the class of *admissible measures* (see Section 3) are measures for which the inverse Cauchy transform (and hence the $R$ or $S$ transforms) can be accurately computed pointwise on an appropriate domain.

The class of *invertible measures*, described next, are a subset of the class of admissible measures consisting of:

- square-root decaying measures: measures supported on an interval $(a, b)$ with square root singularities at both the endpoints (such as the semi-circle),
- smoothly decaying measures: smooth measures supported on the entire real line (such as the Gaussian), and
- half square-root/smoothly decaying measures: supported on an unbounded interval $(a, \infty)$ or $(-\infty, b)$ with a square root singularity at the finite endpoint.

Invertible measures are a class of measures for which we can *recover* the output measure accurately from knowledge of its inverse Cauchy transform. The results of Section 2 state that the result of a free probability operation, if its support is simply connected, is

generically an invertible measure. The utility of the invertible measures can be discerned from Table 2.

The key idea behind the proposed method is that invertible measures that are represented via a Chebyshev or Fourier series expansion as in the second column of Table 2 have Cauchy transforms whose series expansions are closely related, as listed in the third column of Table 2. Thus, given a series truncation, we can efficiently compute the inverse Cauchy transform $G_\mu^{-1}(y_i)$ at $y_i$ by a companion matrix method. This knowledge of the inverse Cauchy transform $G_\mu^{-1}(y_i)$ at points $\{y_i\}_{i=1}^m$ coupled with the relationship (valid for $y$ in the image of $G_\mu$):

$$G_\mu(G_\mu^{-1}(y)) = y,$$

implies that the desired series expansion coefficients $\{\psi_k\}_{i=1}^n$ for the Cauchy transform representation in the third column of Table 2 can be recovered by solving the Vandermonde system defined by:

$$G_\mu(G_\mu^{-1}(y_i)) = y_i \qquad \text{for } i = 1, \ldots, m > n.$$

This yields Algorithm 3 and Algorithm 5 for smoothly-decaying and square-root decaying measures, respectively. Choosing $n$ appropriately yields the desired level of accuracy. Once these expansion coefficients are computed, we recover the measure $\mu$ via the series expansion in the second column of Table 2.

The recognition that the class of invertible measures has a nice series representation for *both* the measure and its Cauchy transform is an important ingredient of the method; this insight originated in [17] and might be of independent interest to free probabilists. Representing the measures via another basis that yields a sparser series representation of the measure but that does not yield a sparse, directly computable and invertible, series representation of Cauchy transform does not lead to an algorithm for computing the inverse Cauchy transform, thereby stalling progress.

The paper is organized as follows. In Section 2, we discuss the analytic properties of the Cauchy transform that we use to construct the numerical method. In Section 3, we describe a numerical approach for the first sub-task, *i.e.*, the calculation the inverse Cauchy transform for several types of admissible measures that arise in practice. We then solve the inverse problem in Section 4: we develop an algorithm to recover an unknown measure based on pointwise evaluation of its inverse Cauchy transform. In each stage, we achieve spectral accuracy, whenever we know the form of the measure. In the remaining sections, we apply this numerical algorithm to free additive, multiplicative and compressive convolution.

## 2. Regularity properties of free convolution and its implication

We think of admissible measures as candidate 'input' measures that we would like convolve using the operations in Table 1. In this viewpoint, invertible measures are the generic 'output' measures that result from the convolution of admissible 'input' measures.

Table 2 lists the class of invertible measures; recall that these are measures that can be recovered accurately from knowledge of their inverse Cauchy transform. In contrast, admissible measures are those for which we can compute the inverse Cauchy transform accurately.

| Type | Measure | Cauchy transform |
|---|---|---|
| Square-Root Decaying (e.g. Semi-Circle) | $d\mu(x) = \psi(x)\dfrac{2\sqrt{x-a}\sqrt{b-x}}{b-a}dx$, $\psi(x) = \sum_{k=0}^{\infty} \psi_k\, U_k(M_{(a,b)}^{-1}(x))$ | $G_\mu(z) = \pi \sum_{k=1}^{\infty} \psi_{k-1} J_+^{-1}(M_{(a,b)}^{-1}(z))^k$, where $J_+^{-1}(z) = z - \sqrt{z^2-1}$ and, $M_{(a,b)}(z) = \frac{a+b}{2} + \frac{b-a}{2}z$ |
| Half Square Root /Smoothly Decaying | $d\mu(x) = \psi(x)\dfrac{2\sqrt{x-a}}{1+x-a}dx$, $\psi(x) = \sum_{k=0}^{\infty} \psi_k\, U_k(M_{(a,\infty)}^{-1}(x))$ | $G_\mu(z) = \pi \sum_{k=1}^{\infty} \psi_{k-1}\left[J_+^{-1}(M_{(a,\infty)}^{-1}(z))^k - 1\right]$, where $J_+^{-1}(z) = z - \sqrt{z^2-1}$ and, $M_{(a,\infty)}(x) = a + \frac{1+x}{1-x}$ |
| Smoothly Decaying (e.g. Gaussian) | $d\mu(x) = \psi(x)dx$, $\psi(x) = \sum_{k=-\infty}^{\infty} \psi_k u(x)^k$, where $\psi_k = \overline{\psi}_{-k}$ and, $u(x) = \dfrac{i-x}{i+x}$ | $G_\mu(z) = -\sum_{k=0}^{\infty}(-1)^k\psi_k + -2\pi\begin{cases}\sum_{k=0}^{\infty} \psi_k\, u(z)^k, \Im(z) > 0 \\ \sum_{k=-1}^{-\infty} \psi_k\, u(z)^k, \Im(z) < 0\end{cases}$, where $u(z) = \dfrac{i-z}{i+z}$ |

TABLE 2. Series representation of invertible measures and their associated Cauchy transforms.

The semi-circle and Gaussian measures are both invertible and admissible; the uniform measure on an interval and the (discrete) point measure are admissible but not invertible. Invertible measures are thus a proper subset of the class of admissible measures. Might this be a shortcoming of our proposed method? We assert otherwise and argue why the mathematics of free convolution gives us license to carve out the smaller class of invertible measures from the larger class of admissible measures.

Simply put, the free convolution of two admissible measures generically results in an invertible measure. An important implication is that we can predict the form of the convolved measure and apply a suitable algorithm (see Section 4) for recovering the measure from its inverse Cauchy transform.

In this paper, we restrict ourselves to measures that have a density; let $\mu_1$ be strictly positive everywhere then for any non-trivial $\mu_2$ we have that $\mu_1 \boxplus \mu_2$ is absolutely continuous with an analytic density supported on the entire real line. This follows directly from the results in [3]. The subordination theory of free convolution implies that the worst behavior of the two measures is preserved in the convolution. These statements carry through for free multiplicative convolution as well. Thus the convolution of admissible, smoothly decaying measures results in an invertible smoothly decaying measure.

Now consider two (non-point) measures; each supported on an interval (that decay 'fast enough' at the edge of the support). Then, by Theorem A.1, the resulting measure will be an invertible square-root decaying measure (thanks to Serban Belinschi for supplying the proof) and supported on an interval (see [24]). A similar argument (omitted here) holds for free multiplicative convolution as well.

The half-square root/smoothly decaying invertible measure arises from the free convolution of a measure supported on an interval with a measure supported on a semi-infinite interval. The argument in Theorem A.1 may be adapted to show square-root decay at one edge of the support.

The setting that is the least well-characterized in this framework is the convolution of two point measures. Here anything can happen. For example, the convolution of $\mu_1(x) = 0.5\delta(x) + 0.5\delta(x-1)$ with $\mu_2(x) = \mu_1(x)$ yields the arc-sine distribution which is neither admissible nor invertible (according to our definition). An improved characterization of the class of measures that results from the free convolution of point measures would help determine if and how the proposed method may be extended to this setting.

## 3. Computation of inverse Cauchy transforms

We consider the numerical calculation of the Cauchy transform and its inverse for four types of measures, which we refer to as *admissible*:

(1) smoothly decaying measures: smooth measures supported on the real line;
(2) square root decaying measures: measures supported on an interval $(a, b)$ with square root singularities at the endpoints;
(3) half square root/smooth decaying measures: measures supported on $(a, \infty)$ or $(-\infty, b)$ with a square root singularity at the finite endpoint;
(4) point measures and
(5) combinations of the above.

For brevity, we omit the details for half square root/smoothly decaying measures below, as they can be treated very similarly to square root decaying measures. We include the relevant formulæ in Table 2.

We note that the formulæ for the Cauchy transforms below follow from Plemelj's lemma [13]: i.e., if $\mathrm{d}\mu = \psi \mathrm{d}x$ for suitably smooth $\psi$, then

$$\phi^+(x) - \phi^-(x) = -2\pi i \psi(x) \text{ and } \phi(\infty) = 0$$

if and only if $\phi = G_\mu$, where $\phi^+$ denotes the limit in the complex plane from above and $\phi^-$ denotes the limit from below.

3.0.1. *Computing the Cauchy transform and its function inverse of smoothly decaying measures.* A *smoothly decaying measure* is of the form

$$\mathrm{d}\mu(x) = \psi(x)\mathrm{d}x,$$

where $\psi \in C^1(-\infty, \infty)$, $\psi'$ has bounded variation and $\psi(x) = \frac{\alpha}{x} + O(x^{-2})$ as $x \to \pm\infty$ for some constant $\alpha$. We can expand

$$\psi(x) = \sum_{k=-\infty}^{\infty} \psi_k \left(\frac{i-x}{i+x}\right)^k, \tag{1}$$

where $\psi_k = \bar{\psi}_{-k}$ (since $\psi$ is real-valued) and $\psi(\infty) = \sum_{k=-\infty}^{\infty}(-)^k\psi_k = 0$. The Cauchy transform satisfies [27, 17]

$$G_\mu(z) = -2\pi i\left[\begin{cases}\sum_{k=0}^{\infty}\psi_k\left(\frac{i-z}{i+z}\right)^k & \Im z > 0\\ -\sum_{k=-\infty}^{-1}\psi_k\left(\frac{i-z}{i+z}\right)^k & \Im z < 0\end{cases} - \sum_{k=0}^{\infty}(-)^k\psi_k\right]. \tag{2}$$

If $\psi$ is $C^\infty(-\infty,\infty)$ and $\psi$ has a full asymptotic expansion that matches at $\pm\infty$, then the series (1) converges spectrally quickly. Moreover, we can rapidly compute the coefficients of the expansion by applying the FFT to the pointwise function samples $\psi\left(i\frac{1-\mathbf{u}_m}{1+\mathbf{u}_m}\right)$, where $\mathbf{u}_m$ are $m$ evenly spaced points on the unit circle:

$$\mathbf{u}_m = \left[-1, e^{i\pi\left(\frac{2}{m}-1\right)}, \ldots, e^{i\pi\left(1-\frac{2}{m}\right)}\right].$$

Thus we take $m = 2n+1$ and approximate

$$G_\mu(z) \approx -2\pi i\left[\begin{cases}\sum_{k=0}^{n}\psi_k\left(\frac{i-z}{i+z}\right)^k & \Im z > 0\\ -\sum_{k=-n}^{-1}\psi_k\left(\frac{i-z}{i+z}\right)^k & \Im z < 0\end{cases} - \sum_{k=0}^{n}(-)^k\psi_k\right].$$

For large $n$, $\psi_k$ are accurate to machine precision.

Now consider the problem of computing $G_\mu^{-1}$. Note that

$$G_\mu\left(i\frac{1-z}{1+z}\right) \approx -2\pi i\left[\begin{cases}\sum_{k=0}^{n}\psi_k z^k & |z| < 0\\ -\sum_{k=-n}^{-1}\psi_k z^k & |z| > 0\end{cases} - \sum_{k=0}^{n}(-)^k\psi_k\right].$$

We can therefore invert the approximation of $G_\mu$ using a companion matrix method. In detail, we compute the eigenvalues $\{\lambda_1^+(y), \ldots, \lambda_n^+(y)\}$ of the matrix

$$\begin{pmatrix} & & & \frac{\psi_0 - \sum_{k=0}^{n}(-)^k\psi_k - \frac{y}{-2\pi i}}{\psi_n}\\ 1 & & & \frac{\psi_1}{\psi_n}\\ & \ddots & & \vdots\\ & & 1 & \frac{\psi_{n-1}}{\psi_n}\end{pmatrix}.$$

Similarly, we compute the eigenvalues $\{\lambda_1^-(y), \ldots, \lambda_n^-(y)\}$ of the matrix

$$\begin{pmatrix} & & & \frac{\psi_{-n}}{\frac{y}{2\pi i} + \sum_{k=0}^{n}(-)^k\psi_k}\\ 1 & & & \frac{\psi_{-1}}{\frac{y}{2\pi i} + \sum_{k=0}^{n}(-)^k\psi_k}\\ & \ddots & & \vdots\\ & & 1 & \frac{\psi_{-1}}{\frac{y}{2\pi i} + \sum_{k=0}^{n}(-)^k\psi_k}\end{pmatrix}.$$

Then

$$G_\mu^{-1}(y) \approx i\frac{1-\lambda(y)}{1+\lambda(y)},$$

where

$$\lambda(y) = \{\lambda_i^+(y) : |\lambda_i^+(y)| \le 1\} \cup \{\lambda_i^-(y) : |\lambda_i^-(y)| \ge 1\}.$$

Because the Cauchy transform is single valued at infinity, there is a neighbourhood of zero such that, for large $n$, $\lambda(y)$ is single valued.

3.0.2. *Computing the Cauchy transform and its function inverse of square root decaying measures.* Suppose that $\mu$ is supported on the interval $(a, b)$ with the form

$$\mathrm{d}\mu(x) = \psi(x)\frac{2\sqrt{x-a}\sqrt{b-x}}{b-a}\mathrm{d}x,$$

where $\psi \in C^1[a, b]$ and $\psi'$ has bounded variation. We can represent

$$\psi(M_{(a,b)}(x)) = \sum_{k=0}^{\infty} \psi_k U_k(x),$$

where $U_k$ denote the Chebyshev polynomials of the second kind and $M_{(a,b)}$ is an affine transformation from the unit interval to $(-1, 1)$:

$$M_{(a,b)}(x) = \frac{a+b}{2} + \frac{b-a}{2}x.$$

Then the Cauchy transform satisfies

$$G_\mu(z) = \pi \sum_{k=1}^{\infty} \psi_{k-1} J_+^{-1}(M_{(a,b)}^{-1}(z))^k, \tag{3}$$

where

$$J_+^{-1}(z) = z - \sqrt{z-1}\sqrt{1+z}$$

is an inverse to the Joukowsky transform

$$J(w) = \frac{1}{2}\left(w + \frac{1}{w}\right).$$

**Remark 1.** We have not found this exact expression for the Cauchy transform of a square root decaying measure in the literature, though directly related expressions are in [17, 18]. In short, it follows from Plemelj's lemma and the fact that

$$\lim_{\epsilon \to^+ 0} i \frac{J_+^{-1}(x+i\epsilon)^k - J_+^{-1}(x-i\epsilon)^k}{2} = U_{k-1}(x)\sqrt{1-x^2},$$

verifiable by the substitution $x = \cos\theta$ [16].

We can compute the coefficients $\psi_k$ whenever we can evaluate $\psi$ pointwise, and thence the Cauchy transform itself. This is accomplished by first computing the expansion in terms of Chebyshev polynomials of the first kind

$$\psi(M_{(a,b)}(x)) \approx \sum_{k=0}^{n-1} \psi_k T_k(x), \tag{4}$$

which can be accomplished by applying the DCT to $\psi(M_{(a,b)}(\mathbf{x}_n))$, where $\mathbf{x}_n$ are $n$ Chebyshev points of the second kind:

$$\mathbf{x}_n = J(\mathbf{u}_{2(n-1)})_{1:n}.$$

We then transform the expansion (4) to a expansion in terms of Chebyshev polynomials of the second kind using the formulæ

$$T_0(x) = U_0(x), T_1(x) = \frac{U_1(x)}{2} \text{ and } T_k(x) = \frac{U_k(x) - U_{k-2}(x)}{2} \text{ for } k = 2, 3, \ldots$$

This approximation will converge spectrally when $\psi \in C^\infty[a, b]$.

3.0.3. *Computing the inverse Cauchy transform of a square root decaying measure.* We want to solve

$$G_\mu(z) = y.$$

Since $J_+^{-1}(J(w)) = w$ for $w$ inside the unit circle, we have

$$G_\mu\left(M_{(a,b)}(J(w))\right) \approx \pi \sum_{k=1}^{n} \psi_{k-1} J_+^{-1}(J(w))^k = \pi \sum_{k=1}^{n} \psi_{k-1} w^k.$$

We can thus solve $G_\mu\left(M_{(a,b)}(J(w))\right) = y$ to find $w(y)$ inside the unit circle using a companion matrix method (as above). Then

$$G_\mu^{-1}(y) = M_{(a,b)}(J(w(y))).$$

3.0.4. *Computing the Cauchy transform and its function inverse of a point measure.* Suppose $\mathrm{d}\mu(x) = \delta(x-a)\mathrm{d}x$. Then its Cauchy transform is trivial:

$$G_\mu(z) = \int \frac{\mathrm{d}\mu}{z-x} = \frac{1}{z-a}.$$

Its inverse is

$$G_\mu^{-1}(y) = \frac{1}{y} + a.$$

3.0.5. *Combination of previous types of measures.* Consider the case where $\mu$ is a sum of point measures, for example, the counting measure

$$\mathrm{d}\mu = \frac{1}{n}\sum_{i=1}^{n} \delta(x-\lambda_i)\mathrm{d}x$$

of one realization of a $n \times n$ random matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. The Cauchy transform can be computed directly using the previous approach, however, its inverse is no longer straightforward to compute. To calculate the inverse, we surround the support of $\mu$ by an ellipse $E_{(a,b),r}$ in the complex plane, on which the Cauchy transform of the measure is smooth. We then exploit analyticity of the Cauchy transform outside of this ellipse.

Define an ellipse $E_{(a,b),r}$ surrounding the interval $(a,b)$ as the image of the unit circle under the map $M_{(a,b)}(J(rw))$, with inverse $\frac{1}{r}J_+^{-1}(M_{(a,b)}^{-1}(z))$. We can then expand a function $g$ defined on $E_{(a,b),r}$ by

$$g(M_{(a,b)}(J(rw))) = \sum_{k=-\infty}^{\infty} g_k w^k, \tag{5}$$

where the coefficients are computable numerically using the FFT as before.

On and outside this ellipse, $G_\mu$ is analytic and vanishes at infinity, therefore $G_\mu(M_{(a,b)}(J(rw))$ is analytic inside the unit circle for $r < 1$ and vanishes at zero. Hence we can efficiently represent it in terms of positive Laurent series:

$$G_\mu(z) = \sum_{k=1}^{\infty} g_k \left[\frac{1}{r} J_+^{-1}(M_{(a,b)}^{-1}(z))\right]^k.$$

Analyticity of this sum implies that the expression holds true for $z$ outside $E_{(a,b),r}$ as well. Mapping this sum back to the union circle allows us to compute $G_\mu^{-1}$ using companion matrix methods.

**Remark 2.** Note that $r$ is a free parameter. As $r$ approaches one, the ellipse approaches the support of the measure. Since $G_\mu$ generically has singularities on the support of $\mu$, the convergence rate of the expansion (5) degenerates. For $r$ small, the ellipse is too large and the region of validity for computing $G_\mu^{-1}$ shrinks. For the numerical examples below, we fix $r$ arbitrarily ($r = .8$). A better approach would be to exploit the connection with the closely related problem of optimizing the radius of circle used in numerical differentiation; a problem solved in [7].

## 4. Recovering a measure from its inverse Cauchy transform

Using the preceding formulæ and the expressions for the transforms below, we can successfully compute the inverse Cauchy transform $G_\mu^{-1}$ of some unknown measure $\mu$ pointwise, which will arise as the output of a free probability operation. If $\mu$ is either a smoothly or a square root decaying measure, we assert that, under broad condtions, the following algorithm will construct an accurate approximation to $\mu$:

**Algorithm 1.** Compute measure from inverse Cauchy transform

Given $G_\mu^{-1}(y)$ which is defined pointwise for $y \in G_\mu(\mathbb{C})$, point cloud $\mathbf{y}_M = (y_1, \ldots, y_M)$ in the upper half plane and the assumed form of the measure $\mu$ (smoothly or square root decaying measure); compute a representation $\mu$ as follows:

**1:** Use Algorithm 2 to prune $\mathbf{y}_M$ so that all points lie inside $G_\mu(\mathbb{C})$;

**2:** If the desired form for $\mu$ is a smoothly decaying measure, use Algorithm 3;

**3:** Otherwise, if the desired form for $\mu$ is square root decaying measure, use Algorithm 4.

The first step of the algorithm is to assure that all sample points lie within $G_\mu(\mathbb{C})$.

**Algorithm 2.** Prune points

Given $G_\mu^{-1}(y)$ which is defined pointwise for $y \in G_\mu(\mathbb{C})$ and point cloud $\mathbf{y}_M = (y_1, \ldots, y_M)$; compute $\mathbf{y}_m \subset G_\mu(\mathbb{C})$ as follows:

**1:** Select the elements of $\mathbf{y}_M$ that satisfy $\operatorname{sgn} \Im y \neq \operatorname{sgn} \Im G_\mu^{-1}(y)$.

**Proposition 4.1.** *Suppose that $\mu$ is smoothly or square root decaying measure. If $G_\mu^{-1}(y)$ can be analytically continued in a domain containing $G_\mu(\mathbb{C})$, then there is a domain $D$ containing $G_\mu(\mathbb{C})$ such that for all $y \in D$, $\Re y \neq 0$, $\operatorname{sgn} \Im G_\mu^{-1}(y) \neq \operatorname{sgn} \Im y$ if and only if $y \in G_\mu(\mathbb{C})$.*

*Proof.* We first show that $G_\mu^{-1}$ has no turning points (points where $(G_\mu^{-1})'$ vanishes) in $G_\mu(\mathbb{C})$ except possibly at $G_\mu(a)$ and $G_\mu(b)$. Note that

$$(G_\mu^{-1})'(y) = \frac{1}{G_\mu'(G_\mu^{-1}(y))}.$$

From (3) the only points where $G_\mu'$ can blow up are at the endpoints $(a, b)$ for square root decaying measures. Similarly, from (2) the only point where $G_\mu'$ can blow up is at infinity for smoothly decaying measure.

Because of the positivity of measures, we know that $\Im G_\mu(z)$ is negative for $\Im z > 0$ and positive for $\Im z < 0$. Therefore, for $y \in G_\mu(\mathbb{C})$, $\operatorname{sgn} \Im G_\mu^{-1}(y) \neq \operatorname{sgn} \Im y$. Moreover, since $G_\mu^{-1}$ is analytic in a domain containing $G_\mu(\mathbb{C})$, and since no point in the complex plane on $G_\mu^\pm(\operatorname{supp}\mu)$ is a turning point, it is true that $\operatorname{sgn} \Im G_\mu^{-1}(y) = \operatorname{sgn} \Im y$ for $y \in D$ outside $G_\mu(\mathbb{C})$.

□

If we assume the measure is smoothly decaying, then we know precisely the form of its Cauchy transform, but we do not know the relevant coefficients of the expansion. The following algorithm computes these coefficients by applying least squares to the equation

$$G_\mu(G_\mu^{-1}(y)) = y,$$

which is valid for $y \in G_\mu(\mathbb{C})$.

**Algorithm 3.** Compute smoothly decaying measure

Given $G_\mu^{-1}$, point cloud $\mathbf{y}_m$ inside $G_\mu(\mathbb{C}) \cap \{z : \Im z > 0\}$ and positive integer $n$; compute a representation of $\mu$ that is smoothly decaying as follows:

**1:** Compute $\psi_k$ by solving the following system in a least squares sense:

$$-2\pi i \left[ \sum_{k=1}^{n} \psi_k \left( \frac{i - G_\mu^{-1}(y_j)}{i + G_\mu^{-1}(y_j)} \right)^k - \sum_{k=1}^{n} (-)^k \psi_k \right] \approx y_j.$$

**2:** Define $\psi_0 = -2\Re \sum_{k=1}^n (-1)^k \psi_k$ and $\psi_{-k} = \bar{\psi}_k$. Then

$$\mathrm{d}\mu \approx \sum_{k=-n}^{n} \psi_k \left( \frac{i - x}{i + x} \right)^k \mathrm{d}x.$$

**Theorem 4.2.** *Suppose that $\mu$ is a smoothly decaying measure, and $\{\mathbf{y}_m\}$ are a sequence of sets of $m$ points lying inside $G_\mu(\mathbb{C}) \cap \{z : \Im z > 0\}$ which cover $G_\mu(\mathbb{C}) \cap \{z : \Im z > 0\}$ as $m \to \infty$ at a sufficiently fast rate (see proof and Appendix B for precise definition). Then there exists $m$ sufficiently large depending on $n$ so that the Algorithm 3 converges to $\mu$ as $n \to \infty$.*

*Proof.* Because of symmetry, including $\bar{\mathbf{y}}_m$ in the least squares system will not alter the approximation of $\mu$. Therefore, denote $[\mathbf{y}_m, \bar{\mathbf{y}}_m] = [y_1, \ldots, y_{2m}]$. Then

$$y_j = G_\mu \left( i \frac{1 - z_j}{1 + z_j} \right)$$

for some (unknown) $z_j$ inside the unit circle. Under this transformation, the least squares system takes the form

$$-2\pi i \sum_{k=1}^{n} \psi_k (z_j^k - (-1)^k) \approx y_j.$$

This is a Vandermonde system, with an unusual distribution of points. However, as $m \to \infty$, the points $z_j$ must cover the unit circle, and therefore convergence follows from Lemma B.2.

□

**Remark 3.** Though we omit the details, it is straightforward to prove that the point sets $\mathbf{y}_M$ that we generate below satisfy the conditions of the preceding theorem, due to the analyticity of the operations involved.

We can adapt this approach to square root decaying measures as well; since, assuming we know the support of the measure, we again know a precise form for its Cauchy transform.

**Algorithm 4.** Compute square root decaying measure

Given $G_\mu^{-1}$, point cloud $\mathbf{y}_m$ inside $G_\mu(\mathbb{C}) \cap \{z : \Im z > 0\}$ and positive integer $n$; compute a representation of $\mu$ that is square root decaying as follows:

**1:** Compute $(a, b) \approx \operatorname{supp} \mu$ using Algorithm 5;

**2:** Compute (real-valued) $\psi_k$ by solving the following system in a least squares sense:

$$\pi \sum_{k=1}^{n} \psi_{k-1} \Re J_+^{-1} \left( M_{(a,b)}^{-1}(G_\mu^{-1}(y_j)) \right)^k \approx \Re y_j \text{ and}$$
$$\pi \sum_{k=1}^{n} \psi_{k-1} \Im J_+^{-1} \left( M_{(a,b)}^{-1}(G_\mu^{-1}(y_j)) \right)^k \approx \Im y_j,$$

where

$$M_{(a,b)}(x) = \frac{a+b}{2} + \frac{b-a}{2} x;$$

**3:** Then

$$\mathrm{d}\mu \approx \frac{2\sqrt{x-a}\sqrt{b-x}}{b-a} \sum_{k=0}^{\infty} \psi_k U_k(M_{(a,b)}(x)) \mathrm{d}x.$$

If $\operatorname{supp} \mu$ is calculated accurately, the convergence of Algorithm 4 follows by the same logic as Theorem 4.2. Thus we are left with one last task: computing $\operatorname{supp} \mu$.

**Algorithm 5.** Compute the support of a square root decaying measure

Given $G_\mu^{-1}$, its first two derivatives and initial guesses $(a_0, b_0)$; compute an interval $(a, b)$ approximating the support of $\mu$ as follows:

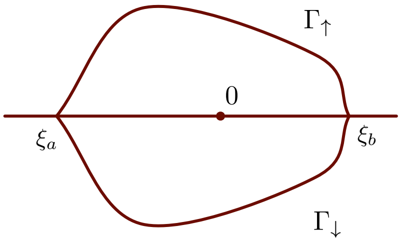


FIGURE 1. A plot depicting the curves on which $G_\mu^{-1}$ is real. These include $\Gamma_\uparrow$, the image of $G_\mu^+$, and $\Gamma_\downarrow$, the image of $G_\mu^-$ and $(\xi_a, \xi_b)$, for $\xi_a = G_\mu(a)$ and $\xi_b = G_\mu(b)$.

**1:** Compute $a$ and $b$ by solving ${G_\mu^{-1}}'(y) = 0$ using Newton iteration, with $a_0$ and $b_0$ as initial guesses.

While we only discussed the computation of $G_\mu^{-1}$, computing its derivative is straightforward since

$$(G_\mu^{-1})'(y) = \frac{1}{G_\mu'(G_\mu^{-1}(y))}$$

and the derived formulæ for Cauchy transforms of admissible measures can be trivially differentiated. Similar logic allows us to compute the second derivative needed to perform the Newton iteration.

The convergence of the algorithm follows:

**Proposition 4.3.** *Suppose $\mu$ is a square root decaying measure. For a sufficiently accurate initial guess, Algorithm 5 will converge to* $\operatorname{supp}\mu$.

*Proof.* Let $[a, b] = \operatorname{supp}\mu$ and

$$\Gamma_\pm = G_\mu^\pm([a, b]),$$

on each of which $G_\mu^{-1}(y)$ is real valued (as its image is $[a, b]$). The Cauchy transform is real valued at the endpoints of the support (since the integrand of the Cauchy transform is integrable), therefore $\Gamma_\pm$ intersect the real axis at $\xi_a = G_\mu(a)$ and $\xi_b = G_\mu(b)$, as depicted in Figure 1. Moreover, $G_\mu^{-1}(y)$ itself is real valued on the real axis (since the Cauchy transform is real valued on $(b, \infty)$ and $(-\infty, a)$). Therefore, $\xi_a$ and $\xi_b$ are saddle points of $G_\mu^{-1}(y)$, and the Newton iteration is guaranteed to converge for sufficiently accurate guesses.

□

## 5. Free additive convolution and the R transform

The *R-transform*, defined as

$$R_\mu(y) := G_\mu^{-1}(y) - 1/y,$$

is the analogue of the logarithm of the Fourier transform for free additive convolution. The free additive convolution of probability measures on the real line is denoted by the symbol $\boxplus$ and can be characterized as follows.

Let $A_n$ and $B_n$ be independent $n \times n$ symmetric (or Hermitian) random matrices that are invariant, in law, by conjugation by any orthogonal (or unitary) matrix. Suppose that, as $n \longrightarrow \infty$, $\mu_{A_n} \longrightarrow \mu_A$ and $\mu_{B_n} \longrightarrow \mu_B$. Then, free probability theory [21, 24] states that $\mu_{A_n+B_n} \longrightarrow \mu_A \boxplus \mu_B$, a probability measure which can be characterized in terms of the $R$-transform as

$$R_{\mu_A \boxplus \mu_B}(y) = R_{\mu_A}(y) + R_{\mu_B}(y). \tag{6}$$

Rearranging (6), we find that

$$G^{-1}_{\mu_A \boxplus \mu_B}(y) = G^{-1}_{\mu_A}(y) + G^{-1}_{\mu_B}(y) - \frac{1}{y}.$$

Therefore, if $\mu_A$ and $\mu_B$ are known in admissible form, we can apply the methods of Section 3 to compute $G^{-1}_{\mu_A \boxplus \mu_B}$. To employ the algorithm of Section 4, we need to assume the form of $\mu_A \boxplus \mu_B$. The discussion in Section 2 allows us to know this *apriori*.

Our last task is to generate a point set $\mathbf{y}_M$ that covers $G_{\mu_A \boxplus \mu_B}(\mathbb{C})$. To accomplish this, we use the following theorem:

**Theorem 5.1.**

$$G_{\mu_A \boxplus \mu_B}(\mathbb{C}) \subset G_{\mu_A}(\mathbb{C}) \cap G_{\mu_B}(\mathbb{C}).$$

*Proof.* This statement first appeared in Proposition 4.3 of [26]. It is a direct consequence of the subordination [6] of the functions expressed in (7). □

Thus we generate $\mathbf{y}_M$ by first generating a point set $\mathbf{z}_{\mu_A,M}$ off the support of $\mu_A$ and then use the point cloud $\mathbf{y}_M = G_{\mu_A}(\mathbf{z}_{\mu_A,M})$. Because we impose that the measures are real, we automatically have that $\bar{G}_\mu(z) = G_\mu(\bar{z})$. Therefore, as required in the algorithms, we restrict $\mathbf{z}_{\mu,M}$ to the upper half plane. Many suitable approaches for generating $\mathbf{z}_{\mu,M}$ exist, the approach we take is the following:

**Algorithm 6.** Generate point sets

Given an a smoothly or square root decaying measure $\mu$; compute a set of points $\mathbf{z}_{\mu,M}$ lying off $\operatorname{supp}\mu$ as follows:

**1:** Generate a point $\mathbf{d}_M$ on the unit disk by taking a tensor product of $\mathbf{u}_m$ with $M^{-1}_{(0,1)}(\mathbf{x}_m)$, the $m$ Chebyshev points on $(0,1)$;

**2:** If $\operatorname{supp}\mu$ is the real line, return $i\frac{1-\mathbf{d}_M}{1+\mathbf{d}_M}$;

**3:** If $\operatorname{supp}\mu$ is an interval $(a,b)$, return the subset of $M_{(a,b)}(J(\mathbf{d}_M))$ which lie in the upper half plane.

### 5.1. Numerical examples.

**Remark 4.** Throughout the paper, we use mean zero and variance $\frac{1}{\sqrt{2}}$ for Gaussian distributions unless otherwise specified. $S_n$ denotes an $n \times n$ random symmetric matrix,

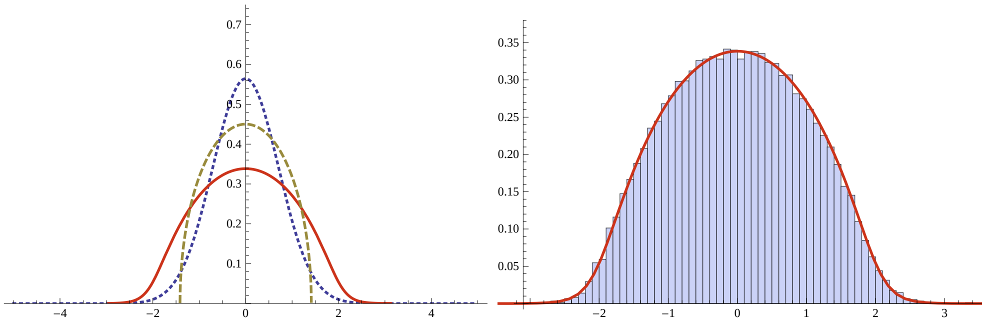


FIGURE 2. Free addition of a Gaussian distribution with a semicircle distribution.

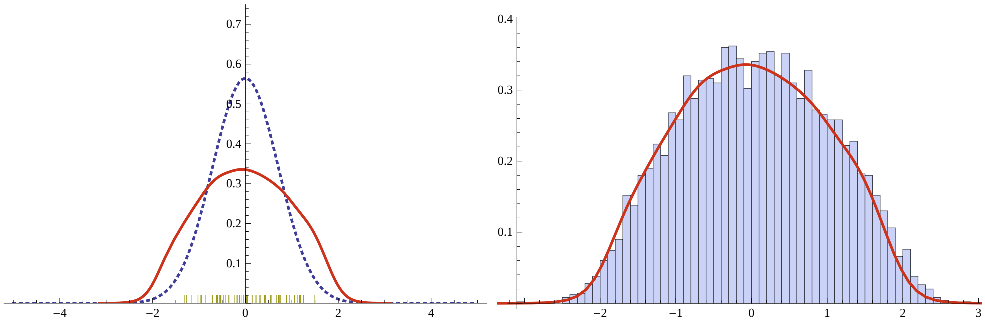


FIGURE 3. Free addition of a Gaussian distribution with a single instance of an approximate semicircle distribution.

constructed by generating a random matrix $A_n$ with Gaussian distributed entries and defining

$$S_n = \frac{A_n + A_n^\top}{\sqrt{2n}}.$$

$Q_n$ denotes a random orthogonal matrix, generated by computing the QR decomposition of $A_n$. Finally, we generate a histogram associated with a random matrix ensemble $B_n$ by computing the eigenvalues of 100 instances of $B_n$.

In Figure 2, we plot the numerically calculated free addition $\mu_G \boxplus \mu_S$ of a Gaussian distribution $\mu_G$ with a semicircle distribution $\mu_S$. This distribution was shown in [8] to be the limiting eigenvalue distribution of a class of Markov matrices. The left graph contains a plot of a Gaussian distribution (dotted), semicircle distribution (dashed) and their free addition (plain). The right graph compares the computed free addition with a histogram of $Q_{150}\Lambda_{150}Q_{150}^\top + S_{150}$, where $\Lambda_n$ is a $n \times n$ diagonal matrix whose entries are Gaussian distributed.

Often one does not have exact expressions for the limiting distributions of the eigenvalues, but rather, one can sample a single instance from the distribution. In this case, the counting measure — a sum of point measures — over this single instance can be calculated. In Figure 3 we repeat the experiment of Figure 2 where the semicircle distribution is replaced with the counting measure $\mu_{A_{50}}$ of a single matrix $A_{50}$ drawn from $S_{50}$. On the right, we compare the computed distribution with the histogram of $Q_{50}\Lambda_{50}Q_{50}^\top + A_{50}$, where $A_{50}$ is now a *fixed* matrix.

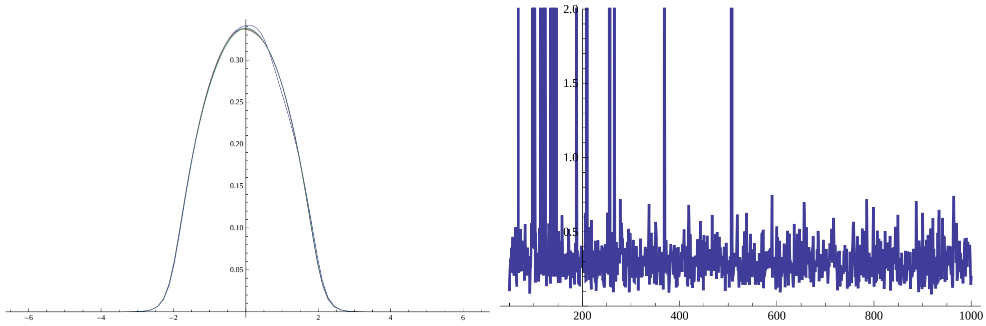


Figure 4. Free addition of a Gaussian distribution with approximate semicircle distributions for $n = 100, 200, \ldots, 400$ (left). The scaled (by $n$) Kolmogorov–Smirnov distance ($D_n = \sup x|F_n(x) - F(x)|$ where $F(x)$ is the distribution in Figure 2 and $F_n$ is the distribution in 3) between the cdfs illustrating convergence in the respective cumulative distribution functions (right).

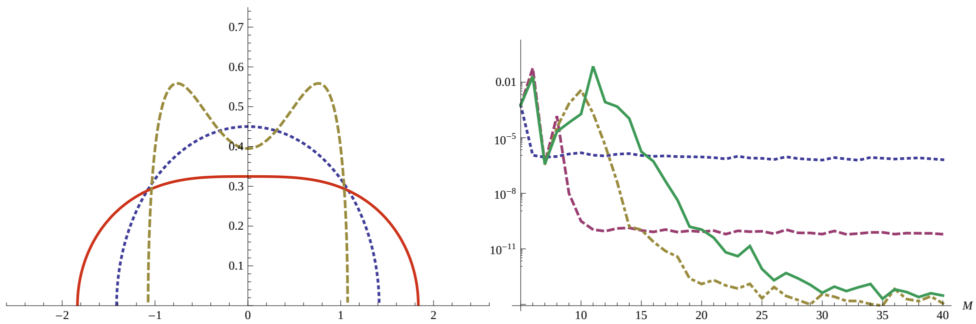


Figure 5. Free addition of a Semicircle distribution with the equilibrium measure associated with the potential $V(x) = x^4$.

As $n \to \infty$, $\mu_{S_n} \boxplus \mu_G$ will converge in some sense to $\mu_S \boxplus \mu_G$, as seen in the right hand of Figure 4. We can estimate this growth by comparing the maximum difference of the cdf of computed measures for growing values of $n$. In the right-hand side of Figure 4, we plot this scaled by $n$, demonstrating that the convergence rate appears to be $O(n^{-1})$.

In Figure 5 we compute a measure which is square root decaying. Here we define $\mu_4$ as the equilibrium measure of the potential $V(x) = x^4$ (see [20] for definition of equilibrium measures), which we know in closed form [10]. We then calculate $\mu_S \boxplus \mu_4$ using Algorithm 1. There is no obvious way of generating a histogram for this measure; hence, unlike other examples, there does not exist a Monte Carlo approach for approximating $\mu_S \boxplus \mu_4$. However, this is an example which was calculated symbolically in [19], hence we can compare our numerically computed measure with the exact measure. We plot the error for $n = 20$ (dotted), 40 (dashed), 60 (dash–dotted) and 80 (plain) as $M$ increases. Recall that $n$ is the number of coefficients in the Chebyshev representation of $\mu_S$ and $\mu_4$ while $M$ is the number of points in the point cloud used in the least-squares based measure recovery algorithm described in Algorithm 6. The error is computed by taking the maximum error over 100 Chebyshev points on the interval $\text{supp}(\mu_S \boxplus \mu_4)$.

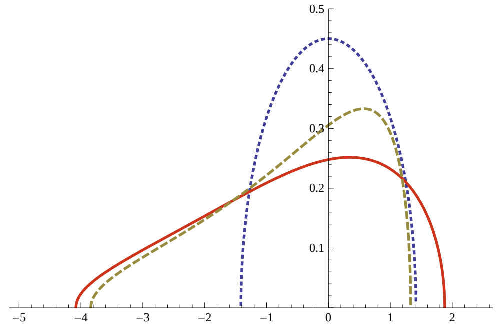


FIGURE 6. Free addition of a Semicircle distribution with the equilibrium measure associated with the potential $V(x) = e^x - x$.

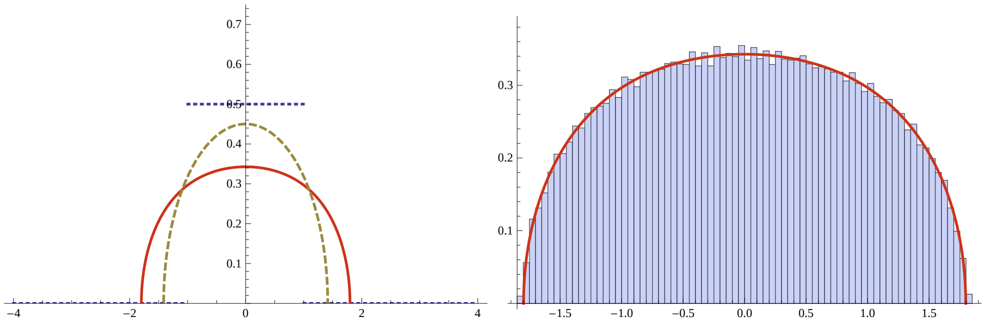


FIGURE 7. Free addition of a step distribution with a semicircle distribution.

In Figure 6, we define $\mu_{EM}$ as the equilibrium measure of the potential $V(x) = e^x - x$ — which we calculate numerically (in the required form) using the approach of [18] — and then calculate $\mu_S \boxplus \mu_{EM}$. This is an example which cannot be computed symbolically, as in [19].

Finally, in Figure 7 we calculate the free addition of a semicircle distribution with a step distribution $\mu_S \boxplus (\frac{1}{2}\mathbf{1}_{(-1,1)})$, by assuming that it is square root decaying, demonstrating that this behaviour is, in fact, generic. While $\frac{1}{2}\mathbf{1}_{(-1,1)}$ does not have the form considered in Section 3, we can in fact calculate its Cauchy transform and inverse Cauchy transform explicitly:

$$G_{\frac{1}{2}\mathbf{1}_{(-1,1)}}(z) = \frac{\log(1+z) - \log(z-1)}{2} \quad \text{and}$$
$$G^{-1}_{\frac{1}{2}\mathbf{1}_{(-1,1)}}(y) = \frac{\coth\frac{y}{2} + \tanh\frac{y}{2}}{2}.$$

Therefore, the algorithms of Section 4 are still usable. We compare the computed distribution with the histogram of $Q_{300}\Lambda_{300}Q_{300}^\top + S_{300}$, where $\Lambda_n$ is a diagonal matrix whose entries are evenly distribution on $(-1, 1)$.

## 6. Free multiplicative convolution and the S transform

In the case where $\mu \neq \delta_0$ and the support of $\mu$ is contained in $[0, +\infty)$, one also defines its *T-transform*

$$T_\mu(z) = \int \frac{x}{z-x} \mathrm{d}\mu(x) \qquad \text{for } z \notin \operatorname{supp} \mu.$$

The *S-transform*, defined as

$$S_\mu(y) := (1+y)/(yT_\mu^{-1}(y)),$$

is the analogue of the Fourier transform for free multiplicative convolution $\boxtimes$. The free multiplicative convolution of two probability measures $\mu_A$ and $\mu_B$ is denoted by the symbols $\boxtimes$ and can be characterized as follows.

Let $A_n$ and $B_n$ be independent $n \times n$ symmetric (or Hermitian) positive-definite random matrices that are invariant, in law, by conjugation by any orthogonal (or unitary) matrix. Suppose that, as $n \longrightarrow \infty$, $\mu_{A_n} \longrightarrow \mu_A$ and $\mu_{B_n} \longrightarrow \mu_B$. Then, free probability theory states [22] that $\mu_{A_n \cdot B_n} \longrightarrow \mu_A \boxtimes \mu_B$, a probability measure which can be characterized in terms of the $S$-transform as

$$S_{\mu_A \boxtimes \mu_B}(z) = S_{\mu_A}(z) S_{\mu_B}(z).$$

The T transform can be computed in the same way as the Cauchy transform; we only need to multiply the function $\psi$ by $x$ beforehand. This is true of its inverse as well. From the relationship of the S transform, we know that

$$T^{-1}_{\mu_A \boxtimes \mu_B}(y) = T^{-1}_{\mu_A}(y) T^{-1}_{\mu_B}(y) \frac{y}{1+y}.$$

Note that $T_{\mu_A \boxtimes \mu_B} = G_{\mu_C}$, for the (non-probability) measure $\mu_C$ defined by

$$\mathrm{d}\mu_C(x) = x \mathrm{d}[\mu_A \boxtimes \mu_B](x).$$

Therefore, we can use the Algorithm 1 to find $\mathrm{d}\mu_C$, and in turn $\mu_A \boxtimes \mu_B$. Similar to free addition, we use the point cloud $\mathbf{y}_M = T_{\mu_A}(\mathbf{z}_{\mu_A, M})$.

**Theorem 6.1.**

$$T_{\mu_A \boxtimes \mu_B}(\mathbb{C}) \subset T_{\mu_A}(\mathbb{C}) \cap T_{\mu_B}(\mathbb{C}).$$

*Proof.* This is a direct consequence of the subordination of the functions expressed via the relationship $T_{\mu_A}(F_A(z)) = T_{\mu_A \boxtimes \mu_B}(z) = T_{\mu_B}(F_B(z))$ derived in [6, Theorem 3.5, pp. 157]. □

In Figure 8, we consider the problem of computing a free product of a a shifted semicircle distribution with a singular Marčenko–Pastur distribution

$$\mathrm{d}\mu_{MP}(x) = \frac{\sqrt{4-x}}{2\pi\sqrt{x}} \mathrm{d}x.$$

While this distribution is not admissible, it is when we multiply by $x$; as in the definition of the T-transform. The procedure then works as before. We compare the computed measure with a histogram of

$$B_{200} B_{200}^\top (S_{200} + 3I),$$

where $B_n = \frac{1}{\sqrt{n}} A_n$ and $A_n$ is an $n \times n$ random matrix with Gaussian distributed entries, now with mean zero and variance one.

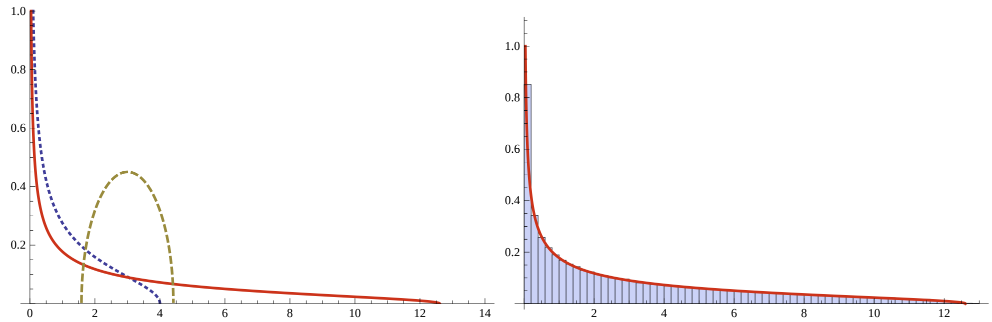


Figure 8. Free times of a shifted semicircle distribution with a Marčenko–Pastur distribution.

## 7. Free compression

Let $B_n$ be the $n \times n$ matrix generated by the taking the upper left $n \times n$ block of $Q_m A_m Q_m^\top$, where $n \leq m$. If the eigenvalues of $A_m$ tend to the distribution $\mu$, then the eigenvalues of $B_n$ tend to the *free compression* of $\mu$, i.e.,

$$\mu_{B_n} \to \frac{m}{n} \boxdot \mu.$$

Let $\alpha \in (0, 1]$. We have that [14]

$$R_{\alpha \boxdot \mu}(z) = R_\mu(\alpha z).$$

Rearranging the definition of the R transform, we find that

$$G^{-1}_{\alpha \boxdot \mu}(y) = G^{-1}_\mu(\alpha y) + \frac{1}{y} - \frac{1}{\alpha y}.$$

Therefore, we can apply Algorithm 1 to compute $\alpha \boxdot \mu$, with the point cloud $\mathbf{y}_M = G_\mu(\mathbf{z}_{\mu,M})$.

### 7.1. **Numerical examples.**

In Figure 9, motivated by the theoretical results in [2], we compare the compute free compression of a Gaussian distribution with a histogram of the $\alpha 300 \times \alpha 300$ principal block of $Q_{300} \Lambda_{300} Q_{300}^\top$, where $\Lambda_n$ is an $n \times n$ diagonal matrix whose entries are Gaussian distributed.

## 8. Extensions

We now identify some extensions of the proposed method:

- Measures supported on multiple intervals. Here there are two issues that must be overcome: computation of the inverse Cauchy transform, and determination of the support of the measure. The major complication is that the inverse Cauchy transform is multi-valued.
- Free rectangular convolution (see [5]). This operation inherently requires computation with measures supported on multiple intervals.

We plan to release a software implementation based on the ideas described in this paper.

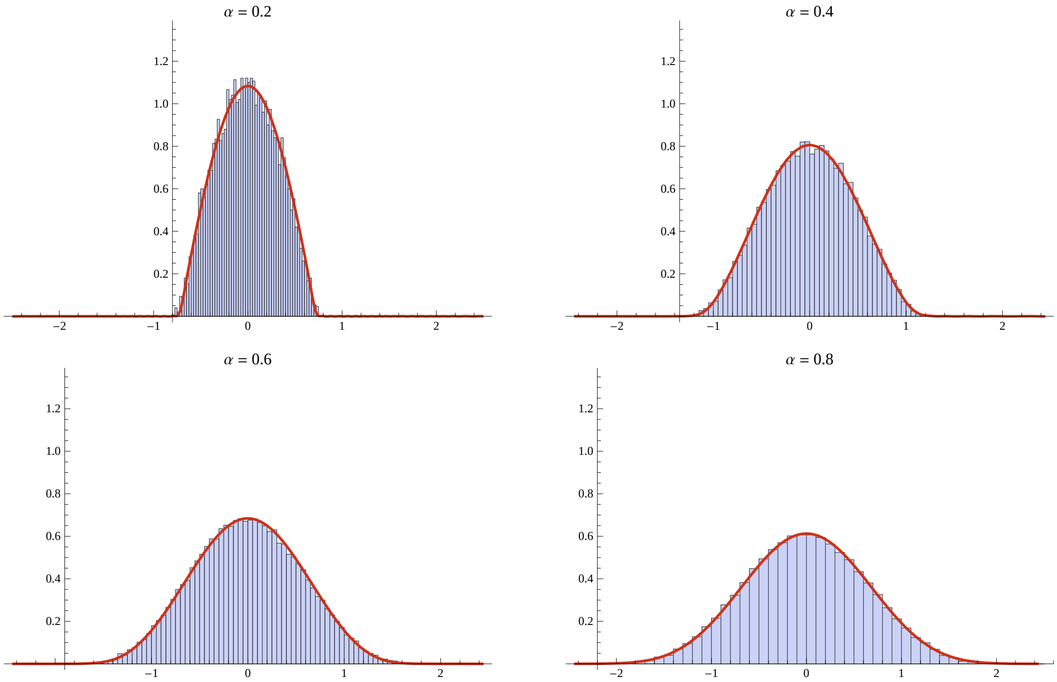


Figure 9. Four compressions of a Gaussian.

## Acknowledgements

We thank Serban Belinschi for many insightful comments regarding the regularity properties of free convolution and for supplying the proof of Theorem A.1. We thank Ben Adcock for suggesting the proof of Lemma B.2. We thank Folkmar Bornemann for initiating this collaboration by pointing R.R.N to S.O's work [17] in response to a query about whether free convolutions might be computable numerically. R.R.N's work was supported by an Office of Naval Research Young Investigator Award N00014-11-1-0660.

## Appendix A. Square-root decay as generic edge behavior of a convolved measure

**Theorem A.1.** *Assume that both $\mu_1$ and $\mu_2$ are compactly supported. If $\mu_1$ and $\mu_2$ do not have 'very fast decay' at the boundary of their respective supports, then $\mu_1 \boxplus \mu_2$ is square-root decaying at the left and right edge of the support.*

*Proof.* We restrict ourselves here to probability measures $\mu_1$ and $\mu_2$ which do not have 'very fast decay' at the boundary of their respective supports. The meaning of 'very fast decay' will be made a bit more precise, but not completely precise, in what follows. Let $\mu_3 = \mu_1 \boxplus \mu_2$ and for $j = 1, 2, 3$, let $F_j(y) = 1/G_{\mu_j}(y)$ denote the reciprocal of the Cauchy transform of the measure $\mu_j$ and let $\varnothing_j$ denote the subordination function:

$$F_1(\varnothing_1(z)) = F_2(\varnothing_2(z)) = F_3(z), \tag{7}$$

where by the results of Bercovici and Belinschi [1], we have that the function $\o_2$ (note $\o_1(z) = F_2(\o_2(z)) - \o_2(z) + z$) satisfies the fixed point equation:

$$\o_2(z) = F_1\left(F_2\left(\o_2(z)\right) - \o_2(z) + z\right) - \left(F_2\left(\o_2(z)\right) - \o_2(z) + z\right) + z. \tag{8}$$

The right hand side of (8) can be expressed as a function of two variables $f(z, w) = F_1(F_2(w) - w + z) - (F_2(w) - w + z) + z$ evaluated at $w = \o_2(z)$. It exhibits good behavior and maps the product of upper half-planes to the upper half-plane. However, what is interesting for our purpose is the fact that $\o_2$ fails to be analytic at a point $x_0 \in \mathbb{R}$ if and only if:

- $F_1'\left(F_2\left(\o_2(x_0)\right) - \o_2(x_0) + x_0\right)\left(F_2'\left(\o_2(x_0)\right) - 1\right) - F_2'\left(\o_2(x_0)\right) = 0$ or,
- one of $F_j$ fails to be analytic at $\o_j(x_0)$.

This second possibility, as can be seen from the first condition (recall that $\o_1(z) = F_2(\o_2(z)) - \o_2(z) + z$), requires that the product $(F_1'(\o_1(x)) - 1)(F_2'(\o_2(x)) - 1)$ stays strictly less than one when $\o_j(x)$ moves all the way to the support of the continuous part of $\mu_j$; this is the sense in which 'the decay at the boundary should not be very fast'.

Now assume that this is the case and that $\left(F_1'\left(\o_1(x)\right) - 1\right)\left(F_2'\left(\o_2(x)\right) - 1\right) = 1$ has a solution at a point $x_0 \in \mathbb{R}$ for which both $\o_j(x_0)$ are real and in the domain of analyticity of $F_j$. Thus $\o_2$ is analytic and we are in a position to apply Weierstrass' preparation theorem (see, for example [9]) to the function $(z, w) \mapsto f(z, w) - w$ around $(z, w) = (x_0, \o_2(x_0))$. The theorem states that:

$$f(z, w) - w = (w^k + c_1(z)w^{k-1} + \cdots + c_k(z))\phi(z, w),$$

where $\phi(z, w)$ is nonzero on a neighborhood of the chosen point, $c_j$ are just analytic maps around $x_0$, and $k$ is the number of zeros with multiplicity of $w \mapsto f(x_0, w) - w$ on some small enough neighborhood.

An application of this theorem reveals that right of the supports of the $\mu_j$'s, we have that $F_j''(w) < 0$, $F_j'(w) > 1$, and we obtain that the second derivative of $w \mapsto f(x_0, w) - w$ is strictly negative at $w = \o_2(x_0)$.

Thus, $k = 2$, and $\o_2$ has a square root singularity at $x_0$. Applying the same argument we can also show that $\o_1$ has a square root singularity at $x_0$, and hence so does $F_3$ since $F_3(z) = F_1(\o_1(z)) = F_2(\o_2(z))$. This implies that $G_{\mu_3}(z)$ will have a square root singularity at $x_0$, the endpoint of the support and hence, by Plemelj's inversion formula, $\mu_3$ will be square-root decaying. The same argument works for the left-hand side of the support as well. □

## Appendix B. Convergence of Vandermonde systems with large number of points

The following proofs are straightforward (we thank Ben Adcock for help proving them), though we have not found them in precisely this form in the literature.

**Proposition B.1.** *Let* $\mathbf{d}_m = (d_1, \ldots, d_m)$ *be a point cloud that covers the unit disk as* $m \to \infty$*, and*

$$V = \begin{pmatrix} 1 & d_1 & \cdots & d_1^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & d_m & \cdots & d_m^{n-1} \end{pmatrix},$$

*the* $m \times n$ *Vandermonde matrix associated with the point cloud. Then for any* $0 < \epsilon$ *and* $n$*, there exists* $m$ *large enough so that* $\|V^+\| \leq \sqrt{n} + \delta$*, where* $V^+$ *denotes the Moore–Penrose pseudoinverse of* $V$*. Furthermore, if, for all* $|x| \leq 1$ *and* $\epsilon > 0$*, the smallest* $m$ *such that* $\min(|x - \mathbf{d}_m|) \leq \epsilon$ *satisfies* $\frac{1}{m} = O(\epsilon^\alpha)$ *for some* $\alpha > 0$*, then* $m = O(n^\alpha)$*.*

*Proof.* Assuming that $V^+$ has full column rank (which will follow from the argument below for large $m$),

$$\|V^+\| = \frac{1}{\sigma_{\min}} = \frac{1}{\inf_{\mathbf{c} \in \mathbb{C}^n, \|c\|=1} \|V\mathbf{c}\|}.$$

where $\sigma_{\min}$ is the smallest singular value. For $m$ large enough, there exist $n$ points within $\frac{1}{n}$ of $n$ evenly spaced points $\mathbf{u}_n$. (Under the secondary hypothesis, this $m$ clearly grows like $O(n^\alpha)$.) Let $V_g$ be the $n \times n$ Vandermonde matrix associated with these points, so that (under a certain ordering)

$$V = \begin{pmatrix} V_g \\ V_b \end{pmatrix}.$$

Then

$$\|V\mathbf{c}\| = \left\| \begin{pmatrix} V_g\mathbf{c} \\ V_b\mathbf{c} \end{pmatrix} \right\| \geq \|V_g\mathbf{c}\|$$

We have

$$V_g = V_u + \frac{1}{n}\Delta,$$

where $V_u$ is the Vandermonde matrix associated with $\mathbf{u}_n$ (i.e., a discrete Fourier transform) and $\|\Delta\| \leq 1$. Thus $\|V_g\mathbf{c}\| = \|V_u\mathbf{c}\| + O(\frac{1}{n})$. We know that

$$\inf_{\mathbf{c} \in \mathbb{C}^n, \|c\|=1} \|V_u\mathbf{c}\| = \frac{1}{\|V_u^{-1}\|} = \frac{1}{\sqrt{n}}$$

which completes the proof.

□

**Lemma B.2.** *Suppose that, for all* $|x| \leq 1$ *and* $\epsilon > 0$*, the smallest* $m$ *such that* $\min(|x - \mathbf{d}_m|) \leq \epsilon$ *satisfies* $\frac{1}{m} = O(\epsilon^\alpha)$ *for some* $\alpha > 0$*. If* $f$ *is analytic in the unit disk, then for* $m$ *large enough the least squares approximation of* $f$ *at the points* $\mathbf{d}_m$ *converges to* $f$*.*

*Proof.* Let

$$P_n = (I_n, \mathbf{0})$$

denote the $n \times \infty$ projection operator and let $E_m$ be the $m \times \infty$ operator defined by

$$E_m f = f(\mathbf{d}_m).$$

Then we are approximating $f$ by

$$\tilde{f} = P_n^\top V^+ E_m f.$$

Furthermore,

$$P_n f = V^+ E_m P_n^\top P_n f.$$

We thus have the error

$$\begin{aligned} f - \tilde{f} &= f - P_n^\top P_n f + P_n^\top V^+ E_m (P_n^\top P_n f - f) \\ &= (I - P_n^\top V^+ E_m)(f - P_n^\top P_n f). \end{aligned}$$

In other words,

$$\|f - \tilde{f}\| \leq (1 + m(n^{1/2} + \epsilon))\|f - P_n^\top P_n f\|.$$

$\|f - P_n^\top P_n\|$ decays exponentially fast for any analytic $f$. The theorem follows since $m$ grows at most algebraically with $n$.

□

Finally, we remark that the least squares system used in Algorithm 3 is not quite Vandermonde; it is actually of the form

$$\sum_{k=1}^{m} \psi_k (z_k^k - (-1)^k) \approx f(z_k)$$

where $f$ vanishes at $-1$. The logic of the preceding proofs still follow. To see this, define the $n$ points $\tilde{\mathbf{u}}_n$ as the points $\mathbf{u}_{n+1}$ with the point $-1$ removed. Interpolating $f$ at $\tilde{\mathbf{u}}_n$ by $(z+1, z^2-1, \ldots, z^n - (-1)^n)$ will also interpolate $f$ at $-1$, hence it is equivalent to interpolating $f$ at $\mathbf{u}_{n+1}$. Thus the norm of the inverse of the relevant interpolation matrix at $\mathbf{u}_n$ is bounded by $\sqrt{n+1}$.

## References


[1] S. T. Belinschi and H. Bercovici. A new approach to subordination results in free probability, *Journal d'Analyse Mathématique*, 101(1): 357–365, 2007.

[2] S. T. Belinschi, M. Bozejko, F. Lehner and R. Speicher. The normal distribution is-infinitely divisible, *Advances in Mathematics*, vol 226, no. 4, pp. 3677-3698, 2011.

[3] S. T.Belinschi. The Lebesgue decomposition of the free additive convolution of two probability distributions, *Probability Theory and Related Fields*, vol 142, no. 1, 125–150, 2008.

[4] S. T.Belinschi. A note on regularity for free convolutions, *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, vol. 42, no. 5, pp 635–648, 2006.

[5] F. Benaych-Georges. Rectangular random matrices, related convolution, *Probability Theory and Related Fields*, vol. 144, no. 3, 471–515, 2009.

[6] P. Biane. Processes with free increments. *Math. Z.*, 227(1):143–174, 1998.

[7] F. Bornemann. Accuracy and stability of computing high-order derivatives of analytic functions by Cauchy integrals. *Found. Comput. Math.* 11: 1–63, 2011.

[8] W. Bryc, A. Dembo and T. Jiang. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *The Annals of Probability*, vol. 34. no. 1, pp. 1–38, 2006.

[9] E. M. Chirka. *Complex Analytic Sets*, vol. 46, Springer, 1989.

[10] P. Deift. *Orthogonal Polynomials and Random Matrices: a Riemann–Hilbert Approach.* American Mathematical Society, 2000.

[11] F. Hiai and D. Petz. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2000.

[12] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.

[13] N. I. Muskhelishvili. *Singular Integral Equations.* Groningen: Noordhoff (based on the second Russian edition published in 1946), 1953.

[14] A. Nica and R. Speicher. On the multiplication of free N-tuples of noncommutative random variables/ *American Journal of Mathematics*, pp. 799–837, 1996.

[15] A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability*, volume 335 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge, 2006.

[16] F. W. J. Olver, D. W. Lozier, R. F. Boisvert and C. W. Clarke. *NIST Handbook of Mathematical Functions* Cambridge University Press, 2010.

[17] S. Olver. Computing the Hilbert transform and its inverse. *Maths Comp.*, 80 (2011):1745–1767.

[18] S. Olver. Computation of equilibrium measures. *J. Approx. Theory*, 163 (2011):1185–1207.

[19] N. R. Rao and A. Edelman. The polynomial method for random matrices. *Found. Comput. Math.*, 8(6), 649–702, 2008.

[20] E. B. Saff and V. Totik. *Logarithmic Potentials with External Fields.* Springer, 1997.

[21] D. Voiculescu Addition of certain non-commuting random variables. *Journal of Functional Analysis*, vol. 66, no. 3, pp. 323–346, 1986.

[22] D. Voiculescu Multiplication of certain non-commuting random variables. *J. Operator Theory*, vol. 18, no. 2, pp. 223–235, 1987.

[23] D. Voiculescu. Limit laws for random matrices and free products *Inventiones Mathematicae*, vol. 104, no. 1, pp. 201–220, 1991.

[24] D. V. Voiculescu, K. J. Dykema, and A. Nica. *Free Random Variables*, volume 1 of *CRM Monograph Series.* American Mathematical Society, Providence, RI, 1992. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.

[25] D. Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory, I. *Communications in Mathematical Physics*, vol. 155(1), pp. 71-92, 1993.

[26] D. Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory, II. *Inventiones Mathematicae*, vol. 118(1), pp. 411-440, 1994.

[27] J. A. C. Weideman. Computing the Hilbert transform on the real line. *Maths Comp.*, 64 (1995):745–762.

[28] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

Raj Rao Nadakuditi, Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Avenue, Ann Arbor, MI 48109. USA.

*E-mail address*: rajnrao@eecs.umich.edu

*URL*: http://www.eecs.umich.edu/~rajnrao/